Year: 2012

Towards continuous monitoring of mental workload

Cinaz, Burcu ; La Marca, Roberto ; Arnrich, Bert ; Tröster, Gerhard

Abstract: Continuous monitoring of mental workload offers new opportunities to support preventing mental disorders and maintaining mental health. In order to achieve a quantification of mental workload, different load levels have to be discriminated. This work goes towards continuous monitoring of mental workload in daily life: we present our experimental designs and the achieved results in discriminating three levels of mental workload in a laboratory simulation and in daily life. In the lab setting we achieved an average recognition rate of 82.14% for all 7 subjects under study. In daily life two subjects were monitored during low, medium and high workload days. We achieved average accuracy rates of 72.03% and 77.66% for both subjects respectively. Finally we have investigated whether the data from the lab experiment are suitable to discriminate low, medium and high workload days.

Originally published at:
Cinaz, Burcu; La Marca, Roberto; Arnrich, Bert; Tröster, Gerhard (2012). Towards continuous monitoring of mental workload. In: 5th International Workshop on Ubiquitous Health and Wellness, Copenhagen, 26 September 2012.

Towards Continuous Monitoring of Mental Workload

Burcu Cinaz, Roberto La Marca, Bert Arnrich, Gerhard Tröster
ETH Zurich, Electronics Laboratory
Gloriastrasse 35, 8092 Zurich, Switzerland
burcu.cinaz@ife.ee.ethz.ch

ABSTRACT

Continuous monitoring of mental workload offers new opportunities to support preventing mental disorders and maintaining mental health. In order to achieve a quantification of mental workload, different load levels have to be discriminated. This work goes towards continuous monitoring of mental workload in daily life: we present our experimental designs and the achieved results in discriminating three levels of mental workload in a laboratory simulation and in daily life. In the lab setting we achieved an average recognition rate of 82.14% for all 7 subjects under study. In daily life two subjects were monitored during low, medium and high workload days. We achieved average accuracy rates of 72.03% and 77.66% for both subjects respectively. Finally we have investigated whether the data from the lab experiment are suitable to discriminate low, medium and high workload days.

Author Keywords

Mental workload simulation, mental workload classification, daily life monitoring, mobile healthcare, heart rate variability

ACM Classification Keywords

J.3. Life and medical sciences: Health; I5.4. Pattern recognition: Applications, I.5.2. Pattern recognition: Design methodology, H.1.2 Models and principles: User/Machine systems

General Terms

Design, Experimentation, Human Factors, Measurement

INTRODUCTION

Recently, the European Foundation for the Improvement of Living and Working Conditions called the attention on work-related stress which was associated with an increasing number of mental disorders [5]. Work-related stress occurs when there is a mismatch between job load and the capabilities of the worker [13]. Since in the developed countries the workplace has changed due to globalization, use of new information and communication technology, mental workload is the dominant element in most jobs. If high level of mental workload cumulates and recovery fails, health problems such as chronic stress, depression or burnout can occur. Continuous monitoring of mental workload offers new opportunities to support preventing mental disorders and maintaining mental health. Most of the existing studies try to discriminate a state of mental load from a resting condition in a laboratory setting. In [2] and [11] two stress factors were investigated under laboratory conditions: high cognitive load under time pressure and social-evaluative threat. In both studies mild cognitive load was discriminated from a constant high stress level. In [12] a mental arithmetic task was used to induce mental workload and the recovery patterns of physiological responses as indicators of stress were investigated.

Continuous monitoring of work-related stress or mental workload is still in an exploratory stage. One example is the research project "Mobile Heart Health", which aims to detect early signs of stress triggered by physiological or contextual changes [10]. In [14] a mobile emotion measurement platform was built and its robustness, usability and usefulness were tested. A recent study presented the results from an experiment aimed at detecting emotional states in daily life [7]. The authors addressed the difficulties of self-reporting and analysis of emotions.

Our work goes towards continuous monitoring of mental workload in daily life. First, we present our results in discriminating different levels of mental workload in a laboratory simulation. Next, we present our approach to monitor mental workload in daily life. Finally we compare the laboratory with the daily life setting. Since for an "everyday life application", a minimal sensor setup is desired for comfort reasons, we employ a single sensor modality: a mobile system to measure heart rate (HR) and body acceleration. The analysis of the heart rate variability (HRV) was chosen, because it represents a sensitive stress and mental load measure. In this work, we investigate HRV features in the time as well as in the frequency domain.

In the following we give an overview about the measurement system. Then we describe the laboratory simulation of different mental workload levels and the daily life scenario. Afterwards we describe the data processing and finally we present and discuss our results achieved from both laboratory and daily life experiment.

Figure 1: Zephyr BioHarness monitoring system

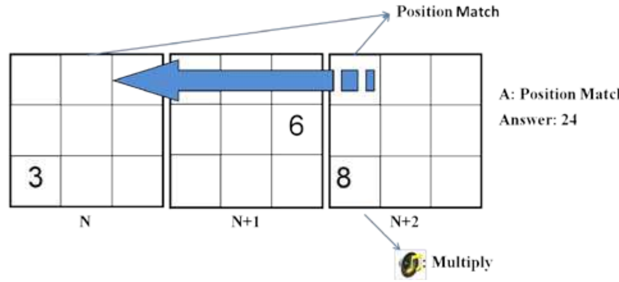

Figure 2: Dual arithmetic 2 back task.

MATERIALS AND METHODS

Mobile ECG Measurement

The physiological responses and body acceleration were measured with the Zephyr BioHarness chest belt as depicted in Figure 1. The monitoring belt consists of three smart fabric sensors to acquire cardiac activity, breathing rate and skin temperature [1]. The ECG data was sampled with 250Hz. The body acceleration is measured with a 3D accelerometer included in the recording device. In addition to ECG data, the chest belt provides RR intervals by measuring the duration between two consecutive R waves of the ECG.

Experiments

In the following we present our experiments. The first experiment addresses how we can simulate and discriminate different levels of mental workload in a laboratory setting. In the second study, we extend the lab experiment with a long term monitoring of daily routines by considering different activities occurring during the day.

Monitoring of Mental Workload in the Lab

Seven healthy subjects participated in this first study (age between 25 and 34 years). Our main goal for this experiment was to simulate different levels of mental workload and to discriminate those levels based on heart rate features obtained from the mobile ECG system. Three sessions each of length 20 minutes with low, medium and high workload were chosen. Subjects performed each session on separate days in the afternoon, while the different sessions were randomly assigned for each subject in order to avoid sequence effects and, therefore, to counterbalance learning effects. We used three variants of the Dual N-Back Task [8] to induce low, medium and high mental workload:

Position 1 Back (Low Workload; very easy task with visual stimuli): A square appears every 4.5 seconds in one of eight different positions on a regular grid on the screen. The subject has to respond by using the keyboard if the position of the currently shown square is the same as the one that was presented just before. This kind of workload is comparable to monotonous monitoring tasks where the subject has to sustain his attention at the same level.

Arithmetic 1 Back (Medium workload; easy task with combined visual and auditory stimuli): An integer number between 0 and 9 appears every 4.5 seconds on the screen. For each number a math operator (add, subtract, multiply or divide) is presented via an audio message. The subject has to apply the math operation on the currently shown number and the one that was presented just before. The result of the calculation has then to be entered on the keyboard. This task reflects medium cognitive load since the subject has to memorize one number and to perform a math task in the given time.

Dual Arithmetic 2 Back (High Workload; demanding task with combined visual and auditory stimuli): In this mode, the two former position and arithmetic tasks are combined. An integer number between 0 and 9 appears every 4.5 seconds in one of eight different positions on a regular grid. For each number a math operator (add, subtract, multiply or divide) is presented via an audio message. The subject has to respond if the position of the currently shown number is the same as the one that was presented 2 positions back. In addition, the subject has to apply the math operation on the currently shown number and the one that appeared 2 positions back. The result of the calculation has then to be entered on the keyboard. An example of this task is shown in Figure 2. This task represents a high cognitive load since the subject has to memorize the position and the value of two numbers and has to perform a math task in the given time.

After each session, subjects were asked to assess their perceived workload. For this subjective rating of perceived workload we employed the NASA Task Load Index (TLX) [6]. In [3] we already have shown that all participants perceived the induced load levels as intended from the experiment design.

Monitoring of Mental Workload in Daily Life

In this study we picked up two subjects who participated in the first experiment and monitored them in three different days which are characterized by their mental demands (low-medium-high). During daily monitoring, participants labeled their activities by taking pictures using a mobile phone. The recording time was between 3.5 and 5 hours. The characterization of low, medium and high workload days are as follows:

Low Workload Day: Subjects were monitored at the weekend which consists of activities such as watching television, internet surfing and going for a walk.

Medium Workload Day: Subjects were monitored during a normal working day. The activities consist of tasks such as programming, reading papers and attending lectures. All activities were performed without any time pressure or social stress.

High Workload Day: Subjects had to perform high mental workload tasks under time pressure and social threat. The following four different tasks had to be solved: programming a difficult task, reading and understanding a research paper, writing a research article and designing a detailed experiment. The subjects had to solve each task within 20 minutes although the actual time requirement of each task was higher. Additionally, after solving each task the subjects had to present the results to an expert specialized in that field in order to induce social stress.

Preprocessing and Feature Extraction

Subjects performed the first experiment in a steady state without any physical activity. On the contrary, the daily monitoring experiment includes different levels of physical activities. Since both physical activity and mental workload lead to variations in heart rate (see Figure 3), we automatically exclude data segments in which physical activity occurred. The procedure is described in the following. In a first step we analyzed the norm of the acceleration signal $|n|=\sqrt{x^2+y^2+z^2}$ during different levels of physical activities. The mean norm was calculated within a time window of one minute with 50% overlapping. In a second step we determined a threshold value of the mean norm which discriminates physical activity from steady states. In the last step we excluded data segments in which the threshold was exceeded. In addition we also excluded three minutes after each physical activity window in order to take the recovery time into account.

For both experimental data sets we removed RR intervals which differ more than 20% from their predecessors in order to remove artifacts. Next, we calculated a set of time and frequency HRV features following the guidelines of the European Task Force [9]: mean heartbeat intervals (Mean RR), standard deviation of RR intervals (SDNN), root mean square of successive differences (RMSSD), the number of successive intervals varying more than 50ms from the previous interval (NN50) and the corresponding percentage (pNN50). In addition, the HRV index (bin width 1/128 sec.), and the triangular interpolation of the R peak interval histogram (TINN) were extracted as geometric parameters. The analysis of HRV features in the frequency domain was done using the Lomb periodogram since it does not require resampling of unevenly sampled signals such as RR data [4]. We used two frequency bands defined as follows: low frequency (LF): 0.04-0.15 Hz and high frequency (HF): 0.15-0.4 Hz. In this work, we used the normalized values of LF, HF and LF/HF bands as the frequency features. Finally, selected features were used for the classification of mental workload levels. Figure 4 illustrates the complete data processing chain comprising the steps of eliminating physical activity segments, RR artifact correction, feature extraction and mental workload classification.

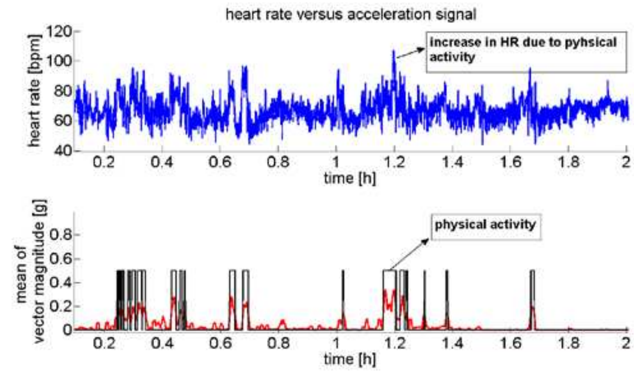

Figure 3: Increase in heart rate due to physical activity

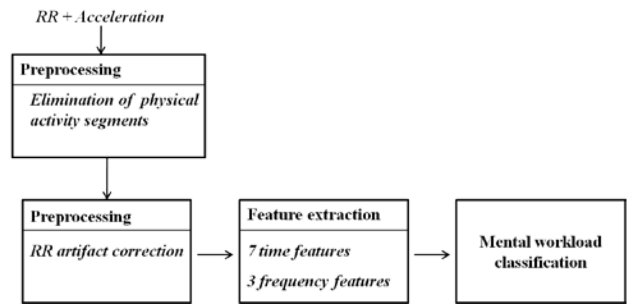

Figure 4: Preprocessing and feature extraction process

DATA ANALYSIS AND RESULTS

Mental Workload Classification of the Lab Experiment

For the first experiment, we divided RR intervals into segments each containing 3 minutes of data with 50% overlapping. In all segments the above mentioned HRV features were computed. For each time window we applied kNN classification using leave-one-out cross validation for each subject individually. The accuracies of the classifier for each subject are shown in Figure 5. For all subjects we achieved an average recognition rate of 82.14% with 12.91% standard deviation.

Mental Workload Classification of Daily Life Experiment

After excluding periods of physical activity, HRV features were extracted for each time window of length 10 minutes. We applied kNN algorithm using ten-fold cross validation to discriminate low, medium and high workload days. Table 1 shows the individual confusion matrices for each subject. The average accuracy rate for the first subject was 72.03% and 77.66% for the second subject. It can be seen that the classification error was highest for the low workload of subject 1. On the contrary, for the second subject low workload day was classified more accurately although high workload had the lowest accuracy.

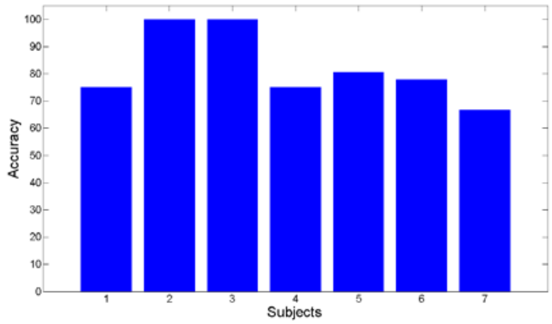

Figure 5: Subject-dependent classification accuracies for discriminating low, medium and high workload

		Actual		
	Subject 1	Low	Medium	High
Predicted	Low	57%	1.9%	5.6%
	Medium	20.6%	84.5%	19.8%
	High	22.4%	13.6%	74.6%

		Actual		
	Subject 2	Low	Medium	High
Predicted	Low	88.9%	2.5%	4.5%
	Medium	7.2%	78.3%	29.9%
	High	3.9%	19.2%	65.6%

Table 1: Mental workload classification of daily life experiment for subject 1 (left) and subject 2 (right)

		Actual		
	Subject 1	Low	Medium	High
Predicted	Low	0%	18%	16%
	Medium	48.2%	44.4%	4.4%
	High	51.7%	37.5%	79.5%

		Actual		
	Subject 2	Low	Medium	High
Predicted	Low	95.4%	89.3%	83.4%
	Medium	4.6%	10.7%	16.1%
	High	0%	0%	0.4%

Table 2: Mental workload classification of daily life experiment based on lab experiment for subject 1 (left) and subject 2 (right)

Mental Workload Classification of Daily Life Experiment based on Lab Experiment

We investigated whether the data from the lab experiment are suitable to discriminate low, medium and high workload days. In the kNN algorithm we used the lab data as training set while the data from daily monitoring served as test set. In Table 2 the individual confusion matrices for subjects 1 and subject 2 are shown respectively. It can be observed that data segments from the high workload day were correctly identified in 79.5% for subject 1. However, the accuracies for correctly identifying the low and medium day are significant lower. In contrast to the first subject, the data segments for the low workload day were correctly identified in 95.4% for subject 2. However, the accuracies for correctly identifying the medium and high day are significant lower.

CONCLUSION

In this paper we presented our experimental results in discriminating different levels of mental workload in a laboratory simulation and in daily life. In the lab setting we achieved an average recognition rate of 82.14% for all 7 subjects under study. In daily life two subjects were monitored during low, medium and high workload days. In a preprocessing step we excluded data segments of physical activity. As a result we achieved average accuracy rates of 72.03% and 77.66% for both subjects respectively. Finally we investigated whether the data from the lab experiment are suitable to discriminate low, medium and high workload days. The results show that the simulated workload in the lab setting does not reflect all workload levels in our daily life scenario.

REFERENCES

[1] Zephyr. http://www.zephyr-technology.com/.

[2] Bert Arnrich, Cornelia Setz, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. What does your chair know about your stress level? *IEEE Transactions on Information Technology in Biomedicine: Affective and Pervasive Computing for Healthcare*, 2010.

[3] Burcu Cinaz, Roberto La Marca, Bert Arnrich, and Gerhard Tröster. Monitoring of mental workload levels. In *Proceedings of IADIS eHealth Conference*, 2010.

[4] Gari D. Clifford. *Signal Processing Methods For Heart Rate Variability Analysis*. PhD thesis, St Cross College, 2002.

[5] European Foundation for the Improvement of Living and Working Conditions. Work-related stress. http://-www.eurofound.europa.eu/, 2007.

[6] S. G. Hart and L. E. Stavenland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, editors, *Human Mental Workload*, chapter 7, pages 139–183. Elsevier, 1988.

[7] Jennifer Healey, Lama Nachman, Sushmita Subramanian, Junaith Shahabdeen, and Margaret Morris. Out of the lab and into the fray: Towards modeling emotion in everyday life. In *Pervasive Computing*, volume 6030/2010, pages 156–173, 2010.

[8] S. M. Jaeggi, M. Buschkuehl, J. Jonides, and W. J. Perrig. Improving fluid intelligence with training on working memory. *Proc. Natl. Acad. Sci. U.S.A.*, 105:6829–6833, May 2008.

[9] Marek Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, and P. J. Schwartz. Heart rate variability: standards of measurement, physiological interpretation and clinical use. *Circulation*, 93:1043–1065, Mar 1996.

[10] Margaret Morris and Farzin Guilak. Mobile heart health: Project highlight. *IEEE Pervasive Computing*, 8(2):57–61, 2009.

[11] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. Discriminating stress from cognitive load using a wearable eda device. *IEEE Transactions on Information Technology in Biomedicine: Personal Health Systems*, 2010.

[12] C. Soga, S. Miyake, and C. Wada. Recovery patterns in the physiological responses of the autonomic nervous system induced by mental workload. In *SICE, 2007 Annual Conference*, pages 1366–1371, Sept. 2007.

[13] G. van Daalen, T.M. Willemsen, K. Sanders, and M.J.P.M. van Veldhoven. Emotional exhaustion and mental health problems among employees doing people work: the impact of job demands, job resources and family-to-work conflict. *Int Arch Occup Environ Health*, 82:291–303, 2009.

[14] Joyce H.D.M. Westerink, Gert-Jan de Vries, Stijn de Waele, Jack van den Eerenbeemd, Marco van Boven, and Martin Ouwerkerk. Emotion measurement platform for daily life situations. In *International Conference on Affective Computing and Intelligent Interaction*, volume 1, pages 704–708, 2009.